

CLAIMS

I claim:

- 1 1. In a vector space representing the latent semantic content of a collection of
2 documents, a method for discerning the presence of at least one sense of a subject term,
3 the method comprising:
4 determining at least one cluster of documents within the vector space, each cluster
5 corresponding to a subset of documents within the vector space containing
6 a subject term, and
7 determining an implicit position within the vector space of at least one sense of
8 the subject term, the implicit position corresponding to at least one
9 determined cluster.
- 1 2. The method in accordance with Claim 1, wherein the vector space is a latent
2 semantic indexed vector space.
- 1 3. In a collection of documents, each document containing a plurality of terms, a
2 method for discerning the presence of at least one sense of a subject term, the method
3 comprising:
4 forming an m by n matrix, where each matrix element (i, j) corresponds to the
5 number of occurrences of term i in document j ;
6 performing singular value decomposition and dimensionality reduction on the
7 matrix to form a latent semantic indexed vector space;
8 determining at least one cluster of documents within the vector space, each cluster
9 corresponding to a subset of documents within the vector space containing
10 a subject term; and
11 determining an implicit position within the vector space of at least one sense of
12 the subject term, the implicit position corresponding to at least one
13 determined cluster.
- 1 4. In a collection of n documents and a reference collection, each document
2 containing terms, the reference collection containing at least one meaning associated with

3 a term, the total number of terms occurring at least once in the document collection equal
4 to at least m , a method for determining a meaning for a sense of a subject term, the
5 subject term found in at least one document and associated with at least one meaning, the
6 method comprising:

7 forming an m by n matrix, where each matrix element (i, j) corresponds to the
8 number of occurrences of term i in document j ;

9 performing singular value decomposition and dimensionality reduction on the
10 matrix to form a latent semantic indexed vector space;

11 determining at least one cluster of documents within the vector space, each cluster
12 corresponding to a subset of the document collection, each member of the
13 subset having at least one occurrence of a subject term;

14 discerning an implicit position of a sense of the subject term, each implicit
15 position corresponding to at least one determined cluster;

16 discerning at least one non-subject term within the vicinity of the implicit position
17 of the sense; and

18 assigning to the sense having a discerned implicit position, the meaning,
19 associated with the term in the reference collection, that correlates best
20 with the discerned non-subject terms closest to the implicit position of the
21 sense.

1 5. In a vector space representing the latent semantic content of a collection of
2 documents, and in a reference collection comprising at least one meaning associated with
3 a term, a method for determining a meaning for a sense of a subject term, the subject term
4 found in at least one document and associated with at least one meaning, the method
5 comprising:

6 determining at least one cluster of documents within the vector space, each cluster
7 corresponding to a subset of the document collection, each member of the
8 subset having at least one occurrence of a subject term;

9 discerning an implicit position of a sense of the subject term, each implicit
10 position corresponding to at least one determined cluster;
11 discerning at least one non-subject term within the vicinity of the implicit position
12 of the sense; and
13 assigning to the sense having a discerned implicit position, the meaning,
14 associated with the term in the reference collection, that correlates best
15 with the discerned non-subject terms closest to the implicit position of the
16 sense.

1 6. The method in accordance with Claim 5, wherein the vector space is a latent
2 semantic indexed vector space.

1 7. In a collection of n documents and a reference collection, each document
2 containing terms, the reference collection containing at least one meaning associated with
3 a term, the total number of terms occurring at least once in the document collection equal
4 to at least m , a method for determining a meaning for an occurrence of a subject term, the
5 subject term found in at least one document and associated with at least one meaning, the
6 method comprising:

7 forming an m by n matrix, where each matrix element (i, j) corresponds to the
8 number of occurrences of term i in document j ;
9 performing singular value decomposition and dimensionality reduction on the
10 matrix to form a latent semantic indexed vector space;
11 discerning the position, within the vector space, of an occurrence of a subject
12 term; and
13 assigning to the occurrence, the meaning, associated with the subject term in the
14 reference collection, that correlates best with non-subject terms closest to
15 the implicit position.

1 8. In a vector space representing the latent semantic content of a collection of
2 documents, and in a reference collection comprising at least one meaning associated with
3 a term, a method for determining a meaning for a occurrence of a subject term, the

4 subject term found in at least one document and associated with at least one meaning, the
5 method comprising:

6 determining at least one cluster of documents within the vector space, each cluster
7 corresponding to a subset of the document collection, each member of the
8 subset having at least one occurrence of a subject term;

9 discerning an implicit position of a sense of the subject term, each implicit
10 position corresponding to at least one determined cluster;

11 discerning at least one non-subject term within the vicinity of the implicit position
12 of the sense; and

13 assigning to the sense having a discerned implicit position, the meaning,
14 associated with the term in the reference collection, that correlates best
15 with the discerned non-subject terms closest to the implicit position of the
16 sense.

1 9. The method in accordance with Claim 8, wherein the vector space is a latent
2 semantic indexed vector space.

1 10. In a collection of n source documents and a collection of x reference documents,
2 each document containing terms, each reference document containing at least one
3 meaning associated with a term, the total number of terms occurring at least once in the
4 combination collections equal to at least m , a method for determining a meaning for a
5 sense of a subject term, the subject term found in at least one source document and
6 associated with at least one meaning, the method comprising:

7 forming an m by $[n + x]$ matrix, where each matrix element (i, j) corresponds to
8 the number of occurrences of term i in document j ;

9 performing singular value decomposition and dimensionality reduction on the
10 matrix to form a latent semantic indexed vector space;

11 determining at least one cluster of documents within the vector space, each cluster
12 corresponding to a subset of the $[n + x]$ documents having at least one
13 occurrence of a subject term;

14 discerning the implicit position of at least one sense of the subject term
15 corresponding to at least one determined cluster; and
16 assigning to at least one sense corresponding to at least one discerned implicit
17 position, the meaning of the subject term closest within the vector space to
18 the implicit position of the sense.

1 11. The method as described in Claim 10 wherein each document in the reference
2 source corresponds to one meaning.

1 12. In a collection of n source documents and a collection of x reference documents,
2 each document containing terms, each reference document containing at least one
3 meaning associated with a term, the total number of terms occurring at least once in the
4 combination collections equal to at least m , a method for determining a meaning for an
5 occurrence of a subject term, the subject term found in at least one source document and
6 associated with at least one meaning, the method comprising:

7 forming an m by $[n + x]$ matrix, where each matrix element (i, j) corresponds to
8 the number of occurrences of term i in document j ;
9 performing singular value decomposition and dimensionality reduction on the
10 matrix to form a latent semantic indexed vector space;
11 discerning the position, within the vector space, of an occurrence of a subject
12 term; and
13 assigning to the occurrence, the meaning, associated with the subject term, closest
14 to the implicit position of the sense.

1 13. The method as described in Claim 12 wherein each document in the reference
2 source corresponds to one meaning.

1 14. In a collection of documents, each document containing a plurality of terms, a
2 computer-implemented method for discerning the presence of at least one sense of a
3 subject term, the method comprising:

4 forming an m by n matrix, where each matrix element (i, j) corresponds to the
5 number of occurrences of term i in document j ;
6 performing singular value decomposition and dimensionality reduction on the
7 matrix to form a latent semantic indexed vector space;
8 determining at least one cluster of documents within the vector space, each cluster
9 corresponding to a subset of documents within the vector space containing
10 a subject term; and
11 determining an implicit position within the vector space of at least one sense of
12 the subject term, the implicit position corresponding to at least one
13 determined cluster.

1 15. In a collection of documents, each document containing a plurality of terms, a
2 computer program product for discerning the presence of at least one sense of a subject
3 term when executed on a computer system, the computer program product comprising:
4 a computer-readable medium;
5 a matrix-forming module stored on the medium that forms an m by n matrix,
6 where each matrix element (i, j) corresponds to the number of occurrences
7 of term i in document j ;
8 a singular value decomposition and dimensionality reduction module stored on
9 the medium and couple to the matrix forming module that forms a latent
10 semantic indexed vector space from the matrix;
11 a clustering module stored on the medium that determines at least one cluster of
12 documents within the vector space, each cluster corresponding to a subset
13 of documents within the vector space containing a subject term; and
14 a sense position determining module stored on the medium an implicit position
15 within the vector space of at least one sense of the subject term, the
16 implicit position corresponding to at least one determined cluster.

- 1 16. In a collection of documents, each document containing a plurality of terms, a
2 computer program product comprising instructions that when executed perform the
3 method comprising the steps of :
- 4 forming an m by n matrix, where each matrix element (i, j) corresponds to the
5 number of occurrences of term i in document j ;
6 performing singular value decomposition and dimensionality reduction on the
7 matrix to form a latent semantic indexed vector space;
8 determining at least one cluster of documents within the vector space, each cluster
9 corresponding to a subset of documents within the vector space containing
10 a subject term; and
11 determining an implicit position within the vector space of at least one sense of
12 the subject term, the implicit position corresponding to at least one
13 determined cluster.